

Review

# Fatigue in Cancer Treatment Studies: Analysis of Placebo Arms

HOLGER HAUCH<sup>1</sup>, BIRTE J. WOLFF<sup>2</sup> and JOHANNES E. WOLFF<sup>3</sup>

<sup>1</sup>*Pediatric Hematology and Oncology, University Children's Hospital Giessen, Hesse, Germany;*

<sup>2</sup>*Swedish Covenant Hospital Department of Gynecology, Chicago, IL, U.S.A.;*

<sup>3</sup>*AbbVie Inc, North Chicago, IL, U.S.A.*

**Abstract.** *Background/Aim: Fatigue and asthenia are common in patients with cancer; and identifying the cause as drug toxicity versus cancer progression is difficult, particularly in clinical trials without control arms. Materials and Methods: We carried out a systematic literature review of fatigue in placebo arms of randomized cancer trials reported in PubMed from 2000 to 2021. Results: Fatigue/asthenia were reported in 100 out of 134 placebo cohorts, and the average of reported frequencies was 22.8%, with a range of 0-83%. Grade 3 or higher fatigue/asthenia was reported in 2.3% (0-17%). Fatigue/asthenia was positively correlated with nausea (R=0.683) Conclusion: For detection of drug toxicity, observations should be flagged when they are higher than the maximum reported in the placebo arm, and the assessment should be supplemented by comparing observations in early oncology trials to literature placebo arms, including both sample sizes and event numbers.*

Fatigue may be the first specific symptom leading to a cancer diagnosis, and often remains a key component of the burden patients carry throughout cancer treatment. The symptom may be caused by a wide variety of mechanisms, including anemia, insomnia, or hypothyroidism. Nonetheless, in many cases of fatigue, the mechanism remains elusive, and for some cancer therapeutics such as lomustine (1) and bromodomain extraterminal domain (BET) inhibitors (2), the symptom may trigger the need for treatment reductions or discontinuation.

For quantitative comparisons, the grading and definition and of fatigue is of the utmost importance. The Common

*Correspondence to:* Dr. med. habil. Johannes E. Wolff, MD, Ph.D., 1709 Napa Suwe Ln, Wauconda, IL 60084, U.S.A. Tel: +1 2242216093, Fax: +1 8479380660, e-mail: johannesevolff@gmail.com

**Key Words:** Meta-analysis, systematic review, adverse events, asthenia, fatigue, placebo, cancer, oncology, randomized trials, CTCAE, review.

Terminology Criteria for Adverse Events (CTCAE) define fatigue as: "A disorder characterized by a state of generalized weakness with a pronounced inability to summon sufficient energy to accomplish daily activities" (3). The descriptions of grades has changed over the years: Version 2 published in 1999 used the Eastern Cooperative Oncology Group (ECOG) status (4), and the Karnofsky and Lansky scales for definition (Table I). Version 3 published in 2006 included a comparison to baseline in grade 2, and defined grade 4 as disabling. The grading in versions 4 and 5 relate the severity to the activities of daily living (Table I).

There are other terms closely related to fatigue: Malaise is defined in CTCAE as: A disorder characterized by a feeling of general discomfort or uneasiness, an out-of-sorts feeling. Lethargy is defined as: A disorder characterized by a decrease in consciousness characterized by mental and physical inertness. Asthenia is not differentiated from fatigue: In CTCAE versions 2 and 3, it is listed together with fatigue with identical definitions and grades; in versions 4 and 5, asthenia is not mentioned.

In general, when reporting AEs in clinical trials, not all symptoms are listed separately. Instead, among closely related terms, the investigators select those that describe the nature of the event best. Thus, an event would typically not be described by both fatigue and asthenia, even when both terms are accurate descriptions.

The frequency of AEs may be compared between treatment and placebo arms of randomized clinical trials to determine if they are caused by the investigational agent. However, in trials that have no control arm, such judgement relies on comparison to an expectation of how often the AE would have occurred without the drug. This study aimed to support these expectations with data focusing on benchmarks for expected frequencies of AEs in oncology trials. The database and technology matured in the process (5-7). For the analysis of fatigue, a broad search of the publication year 2016 was added to the database, and the terms fatigue and asthenia were combined.

Table I. Definitions of fatigue grade by Common Terminology Criteria for Adverse Events (CTCAE) (3).

Grade	CTCAE		
	Version 2	Version 3	Versions 4 and 5
0	None	Not defined for any AE	Not defined for any AE
1	Increased fatigue over baseline, but not altering normal activities	Mild fatigue over baseline	Fatigue relieved by rest
2	Moderate (e.g., decrease in performance status by 1 ECOG level or 20% Karnofsky or Lansky) or causing difficulty performing some activities	Moderate or causing difficulty performing some ADL	Fatigue not relieved by rest, limiting instrumental ADL
3	Severe (e.g., decrease in performance status by $\geq 2$ ECOG levels or 40% Karnofsky or Lansky) or loss of ability to perform some activities)	Severe fatigue interfering with ADL	Fatigue not relieved by rest, limiting self care ADL
4	Bedridden or disabling	Disabling	Not defined for fatigue
5	Not defined for any AE	Not defined for fatigue	Not defined for fatigue

ADL: Activities of daily living; ECOG: Eastern Cooperative Oncology Group.

## Materials and Methods

The data collection integrated results of previous studies. Those included a narrow search among publications from January 2000 to April 2018 (8), supplemented with broad searches for January 2018 to November 2020 (5), January 2020 to March 2021 (6), and 2017 (7). For this analysis, a further broad search for the publication year 2016 was added.

The search algorithm was unchanged from previous work (5, 6), combining three topics by “AND”, namely cancer, randomized, and adverse events, with each of the topics described by its own list of Medical Subject Heading terms combined with “OR”. It resulted in 374 hits for the publication year 2016. Of those, 308 articles were excluded by manual review of the title and bibliographic data. The most common exclusion criteria in this step were (a) trials not with patients with cancer, (b) not randomized trials, (c) the randomized treatments did not include a placebo monotherapy arm, (d) meta-analysis or Cochrane review, and (e) description of trial design without actual data. The remaining 66 publications were reviewed in detail and compared to what was entered already in the database, resulting in the exclusion of a further 44 articles (already included in existing data: 20; placebo treatment-emergent AE table missing: 16; placebo arm was not monotherapy: 12; placebo arm missing: four; not a cancer study: one; too many discrepancies in the AE data: one (9); more than one category of exclusion may apply to individual publications).

The user interface for data entry was an Excel spreadsheet (Microsoft Corporation, Redmond, WA, USA). It included columns to identify the data source and clinical trial (n=7), study type and data selection methods (n=5), demographic and patient population (n=11), treatment (173 drugs), and AEs by reported grade (6,554 columns). At data entry, synonymous AE terms such as “neutropenia” and “neutrophil count decreased” were summarized, and numeric values were converted to the percentage of patients. At this step, fatigue and asthenia were entered separately. The large number of AE columns was in part caused by the 11 different ways the grades of each AE were reported in the medical literature (6, 7).

The data were then exported from Excel into SPSS (SPSS Version 23.0; IBM, Armonk, NY, USA). A cleaning program followed focusing on common spelling errors and adjusting categorical labels in text fields. Filling in missing values started with an automated step of logical deduction – as previously described (5-7). For instance, when the publication reported no AEs leading to death, yet grade 5 data entry was missing from the treatment-emergent AE table, then grade 5 was recorded as 0% for all AEs. Next, the terms fatigue and asthenia were combined into one field “fatigue/asthenia”. When both terms were reported, the higher number was used in the combined field. For further imputation of missing values, linear regression-based models were used to calculate missing values for specific grades based upon other reported grades of the same patient cohort. To determine which pairs of data could be used for such imputation, Pearson correlation coefficients were calculated in all possible combinations, and those with  $R > 0.8$ , and  $p < 10^{-10}$  were implemented. Quality controls followed, and included a manual review of every single datum for 1% of the lines, as well as automated steps ensuring that the sums of variables were consistent (7). By the nature of linear regression, the sums of values for individual grades imputed by various models may not add up to 100%. Discrepancies were evaluated, including source data review and reconsiderations of the imputation algorithms. This led to restricting some of the models by excluding non-predictive outliers, and to the complete elimination of one of the models. The models finally used were: model 1: Grades  $\geq 3$  related to grades 3+4; model 2: grade  $\geq 3$  related to grade 3 (excluding 0% and  $>30\%$ ); model 3: grades 1+2 related to grades  $\geq 1$  excluding 0%; and model 4: grade 4 related to grades 4+5 (excluding grade 4+5 of 0%). Eventually, remaining numeric discrepancies were corrected in the final imputation step by proportionally adjusting the imputed values such that the sums of values were consistent. Finally, validity was assessed comparing aggregated fatigue/asthenia frequencies with other variables, and comparing these findings between raw data and imputed data.

The influence of demographic variables and the relationships between different AEs were assessed in various ways as described previously (5-7). In this process, fatigue/asthenia was first described as a binary variable (yes/no) when any frequency higher than 0% was reported. Quantitative data were then combined for grade 1 and higher, and expressed as a percentage for each published patient cohort. At least two methods were used for each pair of variables. For quantitative variables, Pearson regression (% fatigue *versus* demographic variable) and the SPSS algorithm “compare means” (quantitative variable among patient cohorts with or without fatigue reported) were used. For categorical variables analysis of variance and chi-square tests were used. Additionally, data were presented visually in scatter plots, and box plots. These exploratory analyses were performed for each of the four cut-offs (grade 1 and higher, grade 2 and higher, *etc.*). All analyses and *p*-values were exploratory.

The data selections for the various calculations differed: Raw data without imputation were used for creating the binary terms fatigue/asthenia listed as yes/no. The correlation of grades to each other was derived after logical deduction of missing values. The influence of demographics and relation of AEs to each other was analyzed after maximal imputation and quality control. The calculation of benchmarks used the same data but restricted them to placebo arms only and excluded data from healthy volunteers and cancer-prevention studies. As sensitivity analysis, these calculations were also repeated in the other data selections. Statistical analyses were carried out using SPSS.

The complete method description and the items listed in the Preferred Reporting Items for Systemic Reviews and Meta-Analyses (10) are available from the corresponding author upon request, and include the search algorithm, the complete list of included articles, and the list of synonyms.

## Results

The search retrieved 123 publications describing 134 cohorts of placebo monotherapy with 26,685 individuals. The core analysis to create benchmarks of fatigue/asthenia frequencies was restricted to 92 cohorts and summarizing 19,360 individuals (excluding studies without fatigue data, cancer-prevention studies, and studies with healthy volunteers). The average reported median age was 58 years, 54% were male, 53% had an Eastern Cooperative Oncology Group (ECOG) performance status of 0, and 70% of the studies were phase III. The most common diagnoses were hematologic, colorectal, breast and lung cancer. Eligibility criteria included measurable cancer in 43% of the reported cohorts and newly diagnosed malignancies in 47%. Placebo was given *per os* in 78%. Further details of the demographics are given in Table II.

Fatigue and asthenia were described in various ways: Among placebo-monotherapy controlled randomized trials, 44.8% described fatigue only, 3.5% asthenia only, and 22% both terms separately. Among those that reported both terms, the percentage of individuals reported with fatigue did not correlate with those reported with asthenia ( $R=0.012$ ,  $p=0.903$ ,  $N=105$ ). Neither of the terms was reported in 25.9% of the patient cohorts, and 3.8% used a combined term of fatigue/asthenia. This example was followed for the further

data described here: Fatigue and asthenia were considered synonymous with respect to their grading and frequency – consistent with CTCAE versions 2 and 3. The variable was named “fatigue/asthenia” and it was assigned the value reported in the source regardless of whether the data source listed it as “fatigue” or “asthenia” or “fatigue or asthenia”. The multistep imputation (7) resulted in more data available for analyses. For instance, in raw data, a higher frequency of fatigue was detected in treatment *versus* placebo arms (27.4%  $N=67$  *versus* 21%  $N=59$ , chi-square test  $p=0.017$ ) when analyzing fatigue grade 1 and higher. The equivalent comparison among the imputed variables showed the same phenomenon but with overall higher fatigue frequency, and based upon a larger number of data (treatment 29%  $N=100$ , *versus* placebo 22.8%  $N=92$ ,  $p=0.014$ ). Furthermore, the imputation also allowed similar analyses for other grade cut-offs such as grade 3 or higher (3.4% *vs.* 2.3%,  $p=0.007$ ), while the same comparison in raw data only showed a trend but no significance (3.7% *vs.* 2.3%  $p=0.13$ ).

The influence of demographics on fatigue/asthenia reporting among placebo arms was minimal for most variables. The exception to this was the variable “cancer organ”, which grouped treatment indications by primary cancer type (Figure 1): The probability of differences for grade  $\geq 1$ , 2, 3, and 4 were 0.04, 0.02, 0.03, and 0.21, respectively, in analysis of variance tables. There were no differences in grade 5 (death), since all those values were 0. Among the cancer organ categories, “no cancer organ” ( $n=4$ ) had the lowest frequency with no fatigue/asthenia reported (0% of patients). This category “no cancer organ” included publications with healthy volunteers, and cancer-prevention studies. Benchmarks for individual organ systems were calculated considering eight patient cohorts as the minimum number for calculating meaningful averages. This resulted in specific benchmarks calculated as mean frequency ( $\pm$  standard deviation) of fatigue/asthenia (all grades) for the following organ systems: Liver (26.1 $\pm$ 7.7%,  $N=10$ ), colorectal (23.2 $\pm$ 20%,  $N=10$ ), breast (21.7 $\pm$ 27.7%,  $N=8$ ), hematological (18.9 $\pm$ 9.3%,  $N=13$ ), and lung (17.7 $\pm$ 9.0%,  $N=8$ ). None of the other demographics significantly influenced fatigue among placebo arms. This included ECOG performance status, gender, median age, relapsed refractory *versus* newly diagnosed cancer, previous lines of treatment, measurable cancer, route of medication, study phase, year of publication, and CTCAE version. For some of these variables, the findings from placebo arms differed from those of the treatment arms: Among treatment arms, cases of newly diagnosed cancer were reported to have less fatigue, intravenous therapy application was linked to a higher rate of fatigue than *per os*, the frequency of ECOG 0 correlated negatively with grade  $\geq 3$  fatigue/asthenia ( $N=279$ :  $R=-0.24$ ,  $p=0.000052$ ), and use of later CTCAE version led to reports of lower frequency of fatigue.

Table II. Demographics of study patients.

Variable		Placebo arms (134 cohorts)	Placebo arms selected for benchmarks (92 cohorts)	Treatment arms (152 cohorts)
Number of patients	Mean (range)	199.1 (7-2,253)	192.2 (10-1,933)	238.5 (5-2,256)
Year		2018 (2009-2021)	2018 (2009-2021)	2018 (2009-2021)
Phase		n=134	n=92	n=152
	1	3%	0%	6.6%
	2	29.1%	29.3%	28.3%
	3	67.9%	70.7%	65.1%
Age, years	Median (range)	57.2, 7.1-74, n=129	58.1, 7.1-74, n=88	56.1, 9.5-74.5, n=146
Male, %	Mean (range)	54.7%, 0-100%, n=125	54.4%, 0-100%, n=84	55.5%, 0-100%, n=142
ECOG 0, %	Mean (range)	55.4%, 0-100%, n=93	53.2%, 0-100%, n=62	58.3%, 0-100%, n=102
Lines of previous therapy, n	Median (range)	0.9, 0-4, n=92	1, 0-4, n=65	1, 0-4, n=103
CTCAE version, %		n=96	n=72	n=106
	2	1%	1.4%	0.9%
	3	25%	26.4%	25.5%
	4	71.9%	69.4%	71.7%
	5	2.1%	2.8%	1.9%
Cancer diagnosis, n (%)	Colorectal	18 (13.4%)	10 (10.9%)	20 (13.2%)
	Hematological	16 (11.9%)	13 (14.1%)	16 (10.5%)
	Liver	12 (9%)	10 (10.9%)	14 (9.2%)
	Breast	11 (8.2%)	7 (7.6%)	13 (8.6%)
	Lung	11 (8.2%)	8 (8.4%)	11 (7.2%)
	Prostate	10 (7.5%)	7 (7.6%)	10 (6.6%)
	Ovary	9 (6.7%)	5 (5.4%)	9 (5.9%)
	Soft tissue	8 (6.0%)	7 (7.6%)	8 (5.3%)
	Skin	6 (4.5%)	5 (5.4%)	6 (3.9%)
	Gastric	5 (3.7%)	4 (4.3%)	5 (3.3%)
	Kidney	5 (3.7%)	3 (3.3%)	8 (5.3%)
	No organ system	4 (3.0%)	0 (0%)	10 (6.6%)
	Brain	3 (2.2%)	1 (1.1%)	4 (2.6%)
	Bladder	2 (1.5%)	2 (2.2%)	2 (1.3%)
	Bone	2 (1.5%)	1 (1.1%)	2 (1.3%)
	GIST	2 (1.5%)	2 (2.2%)	2 (1.3%)
	Pancreas	2 (1.5%)	2 (2.2%)	2 (1.3%)
	Uterus	2 (1.5%)	2 (2.2%)	2 (1.3%)
	Esophagus	2 (1.5%)	1 (1.1%)	2 (1.3%)
	Thyroid	2 (1.5%)	2 (2.2%)	2 (1.3%)
	Other endocrine	1 (0.7%)	0 (0%)	3 (2.0%)
	Head and neck	1 (0.7%)	0 (0%)	1 (0.7%)
Tumor status	No measurable cancer	37 (27.6%)	23 (25%)	46 (30.3%)
	Measurable	51 (38.1%)	40 (43.5%)	55 (36%)
	Mixed tumor status	45 (33.6%)	28 (30.4%)	48 (31.6%)
	Newly diagnosed	53 (39.6%)	43 (46.7%)	58 (38.2%)
	R/R	61 (45.5%)	41 (44.6%)	66 (43.4%)
	Mixed new-R/R	13 (9.7%)	8 (8.7%)	13 (8.6%)
	Healthy volunteers	5 (3.7%)	0 (0.0%)	11 (7.2%)
	Prevention	2 (1.5%)	0 (0.0%)	4 (2.6%)
	Not provided	1 (0.7%)	1 (1.1%)	3 (2.1%)
Route of medication	Oral	104 (77.6%)	72 (78.3%)	112 (73.7%)
	IV	19 (14.2%)	15 (16.2%)	27 (17.8%)
	SC	6 (4.5%)	1 (1.1%)	8 (3.9%)
	IM	2 (1.5%)	2 (2.2%)	4 (2.6%)
	ID	2 (1.5%)	2 (2.2%)	2 (1.3%)
	IV+SC	1 (0.7%)	0 (0.0%)	1 (0.7%)

Placebo arms: Includes all cohorts treated with placebo monotherapy regardless availability of fatigue data; Placebo arms selected for benchmarks: excludes healthy volunteers, and patients cohorts with neither fatigue nor asthenia described. Hematological: leukemia/lymphoma; No organ system: studies enrolling regardless of cancer diagnosis and studies with health volunteers; No measurable cancer: studies enrolling only patients in remissions; Measurable: studies enrolling only patients with measurable disease; mixed status: studies enrolling patients regardless of remission status; R/R: relapsed or refractory malignancy; mixed new-R/R: studies enrolling both patients with newly diagnosed and relapsed or refractory disease; Prevention: cancer prevention studies enrolling patients at risk but without cancer diagnosis; IV: intravenous; SC: subcutaneous; IM: intramuscular; ID: intradermal.

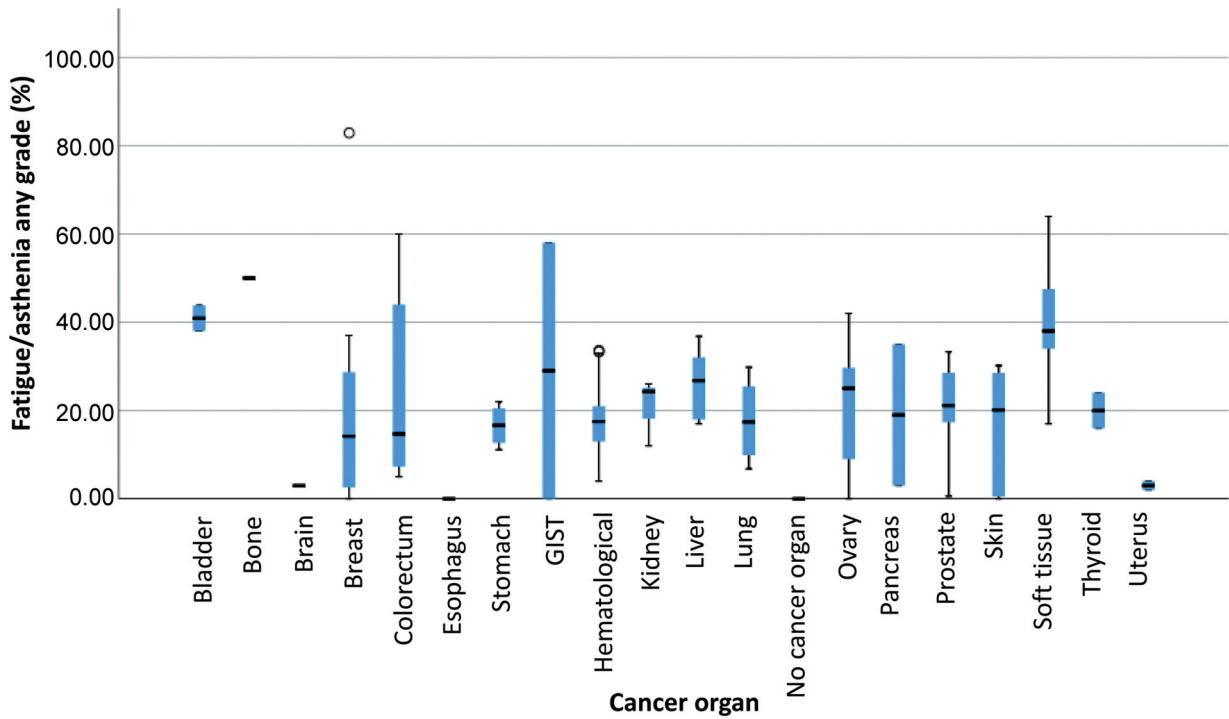


Figure 1. Fatigue/asthenia reported according to cancer organ. Data were restricted to placebo arms of randomized trials with no other cancer therapeutic. Asthenia and fatigue are combined as one variable, regardless of which term was used in reporting. GIST: Gastrointestinal stroma tumors.

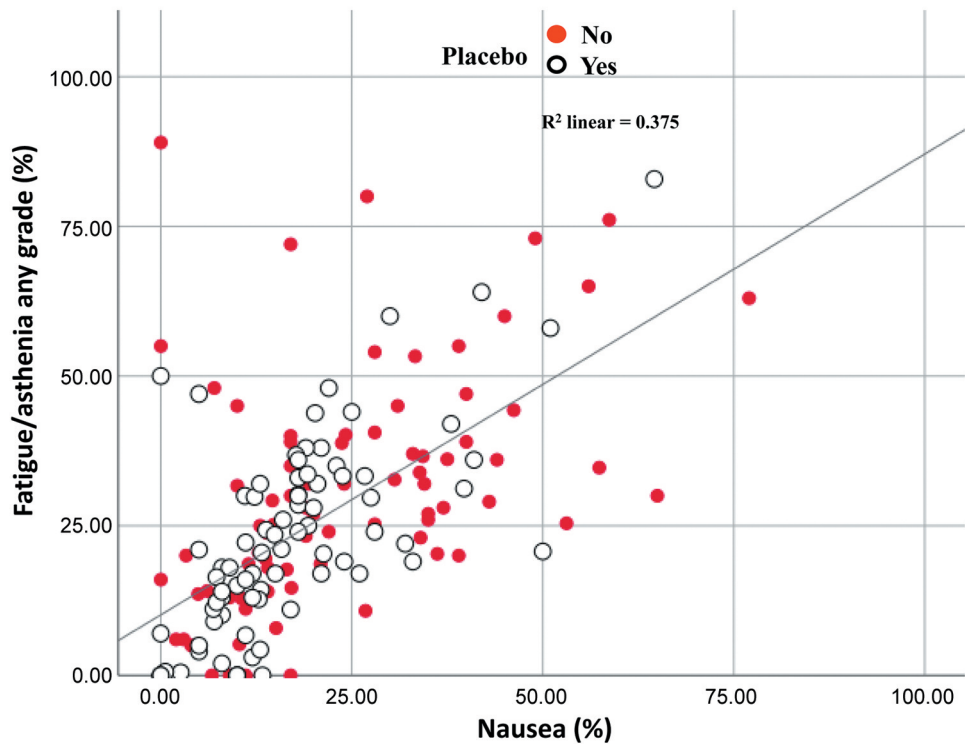


Figure 2. Correlation of nausea with fatigue/asthenia reported in randomized clinical trials. Asthenia and fatigue were combined as one variable, regardless of which term was used in reporting. Percentage values reflect the sum of all grades (grade 1 or higher). The line of best fit was plotted for the total data. The correlation between nausea and fatigue/asthenia was significant among all data, and among subgroups of both treatment and placebo groups.



When relating fatigue/asthenia to other AEs, no negative correlations were found, *i.e.*, in none of the comparisons was a higher rate of fatigue/asthenia related to a lower rate of another AE. In contrast, there were several significant positive correlations. The strongest among those were with nausea, diarrhea, reduced appetite, and anemia. For instance, grade  $\geq 1$  fatigue/asthenia was correlated with grade  $\geq 1$  nausea with  $R=0.683$  ( $N=81$ ,  $p=2.1 \times 10^{-12}$ ; Figure 2), and the equivalent for grade 3 and higher resulted in  $R=0.463$  ( $N=75$ ,  $p=0.000029$ ). There was also a positive correlation with insomnia when grade 3 or higher was considered ( $N=31$ :  $R=0.38$ ,  $p=0.034$ ) but only a trend for grade 1 or higher insomnia ( $p=0.065$ ). Similarly, grade 3 and higher fatigue/asthenia were positively correlated with the frequencies of severe AEs ( $R=0.696$ ,  $N=54$ ,  $p=5 \times 10^{-9}$ ) and discontinuation for drug toxicity ( $N=75$ :  $R=0.359$ ,  $p=0.002$ ), while the equivalent analyses for grade 1 and higher remained insignificant.

For the final analysis to create benchmarks for cancer studies without a control arm, publications with healthy volunteers and cancer-prevention studies were excluded. Among 92 patient cohorts treated with placebo monotherapy, the average reported frequency of grade 1 or higher fatigue/asthenia was  $22.8 \pm 15.9\%$  (Table III). This is the average of the averages, giving relatively more weight to patients in smaller studies. An alternative way to determine the overall frequency is derived from the sum of individual patient numbers. Among placebo arms that described fatigue/asthenia, grade 1 or higher was documented for 3,742 out of 17,679 patients (21.17%). The frequencies calculated in the same way for the other cut-offs were: Grade 2 or higher: 3.1%, grade 3 or higher: 1.7%, grade 4 or higher: 0.04%, and grade 5: 0%, respectively. Further details are given in Table III.

## Discussion

This study combined the data of placebo arms of randomized clinical trials of patients with cancer of any age, and determined the overall frequency of fatigue/asthenia to be 21.2%.

Placebo is the Latin word for “I will please”, describing the therapeutic effect of the physician’s expectation transmitted to the patient’s observation and reporting, and it is considered positive: the placebo effect typically reduces the reported symptom burden. The placebo effect is more likely to occur for AEs that mostly rely on the patient’s sensation, such as fatigue. However, in clinical trials, placebo is an actual medication, which might indeed have adverse effects, even though there is no active agent. Injections cause pain and possibly infection even when the drug itself does not. Niwa and colleagues described a case of severe fatigue caused by an oral medication that was thought to be an inactive placebo, but the capsule included caprylocaproyl polyoxyglyceride,

now suspected to be the causative agent of the fatigue (11). The data provided here do not allow determination of whether a placebo reduces or increases symptoms. However, for the purpose of determining baseline data to judge observations made in drug-development studies, such findings are irrelevant. In any case, data derived from placebo arms are a better data source than those without placebo: Preparations with active agents also have placebo effects, and in the goal to determine the potential causality of the active agent, the placebo effect is an unwanted influencing factor. Using study data from placebo arms as comparator allows some degree of control over the undesired data effect.

Fatigue and asthenia have an overlapping meaning. The Medical Dictionary for Regulatory Activities MEDRA lists both as ‘preferred terms’ under the high-level term “Asthenic conditions”, the high-level group term “General system disorders” and the System Organ Class “General disorders and administration site conditions”. The word “fatigue” is more commonly used, and not just in medicine: In mechanical engineering it describes weakness of metals occurring after repeated mechanical stress. When describing symptoms of a patient with cancer, typically the patient might be described by both “fatigue” and “asthenia” or none of the terms. The overlapping meaning makes it unnecessary to use both in documentation or to analyze them separately when both terms had been used anyway. In fact, in aggregating AE data as percentage of patients, the use of two almost synonymous terms will result in inaccurate, low values for each of the terms. We found the two terms not to be correlated, which further confirms that investigators typically pick one of them. In summary, our findings create doubts in the wisdom of listing both words as ‘preferred terms’. For the purpose of using frequencies to infer drug causality in the context of oncology drug development, we recommended combining the two terms as one quantitative variable.

Fatigue/asthenia can be treated when a somatic cause such as the tumor itself, anemia, hypothyroidism or other hormone disorder is known. Treatment of fatigue/asthenia becomes more difficult in the absence of a treatable cause. Among patients with primary brain tumor, a literature analysis found insufficient evidence for any pharmacological or non-pharmacological treatments for fatigue (12). In contrast, fatigue in menopause was reported treatable with armodafinil (13), and among patients with cancer, exercise and behavioral therapy was reportedly effective (14). Similarly, cognitive behavioral therapy for insomnia was effective in cancer survivors (15). Our data support the concept of treating insomnia to address fatigue, since we found a correlation between the frequency of insomnia and grade 3 and higher fatigue. Furthermore, we found an even closer correlation between fatigue and nausea, which may prompt the hypothesis that antiemetics might also be effective against fatigue in the context of cancer.

Table III. Frequency of fatigue/asthenia among cohorts selected for benchmark calculations aggregated in two different ways.

Grade of fatigue/ asthenia	Placebo			Treatment		
	No. of cohorts	Study arms: Mean±SD (range)	Patient count (%)	No. of cohorts	Study arms: Mean±SD (range)	Patient count (%)
1 or higher	92	22.8±15.9% (0-83%)	3,742 of 17,679 (21.2%)	100	29±18.9% (0-89%),	6,438 of 24,403 (26.4%)
2 or higher	31	3.3±4.9% (0-19%)	171 of 5,509 (3.1%)	37	4.2±8.2% (0-48%)	182 of 6,417 (2.8%)
3 or higher	103	2.3±2.9% (0-17%)	329 of 19,189 (1.7%)	113	3.4±3.4% (0-18%)	925 of 27,638 (3.3%)
4 or higher	77	0.09±0.29% (0-1.6%)	7 of 1,5201 (0.04%)	82	0.2±0.44% (0-2.2%)	45 of 22,800 (0.2%)
5	127	0±0% (0-0%)	0 of 24,405 (0%)	137	0±0.02% (0-0.2%)	1 of 33,759 (0.003%)

Study arms describe aggregation of averages as reported, not taking the cohort size (patient numbers) into account. This gives higher relative weight to patients treated in smaller studies but allows the calculation of range and standard deviations. The patient counts aggregate the absolute patient numbers, resulting in a single frequency for all publications.

Matching of influencing variables is a key component when comparing frequencies in aggregated data. Typically, the assumption is that a tumor response is dependent on the molecular profile of the tumor cells, while AEs are more dependent on demographic variables such as age and comorbidity (5). We found a relevant influence of cancer diagnosis on reporting of fatigue/asthenia. However, other demographic variables remained surprisingly insignificant. Of note, comorbidity may be considered the most influential covariant for AE frequencies, and fatigue might be influenced by various psychological factors. Unfortunately, it was not possible to extract data from the medical literature to address the hypothesis that comorbidities influence fatigue. Thus, the most appropriate benchmarks to be generated from these data may take cancer diagnosis into account, as far as sufficient data are available, but none of the other covariates can be included.

Drug causality is assessed in various workflows. Signal detection often relies on thresholds that can easily be implemented by automated processes monitoring large databases. Any frequency of fatigue/asthenia that is higher than the highest ever reported in placebo arms should be considered drug-related, and flagged. These benchmarks are: 83%, 19%, 17%, and 1% for grade  $\geq 1$ ,  $\geq 2$ ,  $\geq 3$ , and  $\geq 4$  (Table III). Unfortunately, the ranges of reported frequencies were very large, limiting the usefulness of these upper limits of normal. A more sophisticated approach would be to use the actual numbers of patients reported in placebo arms (Table III) and compare them to the study observations, implementing exploratory chi-square tests or more advanced mathematical methods.

Various limitations of aggregating data from published AE tables have been observed over this series of analyses (5-7). Simpson's paradox is the phenomenon in which a trend appears in several groups of data but disappears or reverses when the groups are combined (16). It can be avoided when confounding variables and causal relations are appropriately addressed. We found cancer diagnosis to be a significant

covariant, enabling corrections for this variable. Other interesting observations were made which may also appear as limitations for certain interpretations. Most of these can be summarized as being the result of the information flow, starting with reporting of the AE by the patient, followed by grading and data entry by the local investigator, and after further steps ending in publishing customs: We showed more reporting of headache in healthier patients (5) and a lack of correlation between fatigue and asthenia. Both findings are likely caused by systematically unequal reporting: headache might not be worth mentioning in patients with cancer who have numerous, more severe, symptoms, and asthenia is not entered when fatigue is already documented. Moreover, the degree of reporting diligence varies widely – publications with thorough AE reporting have higher frequencies for all AEs, and combining these data with those generated less diligently will create artificially positive correlations, we even found correlations for opposing medical concepts, such as diarrhea and constipation (7). None of these findings reflect actual biological or medical phenomena. However, they accurately reflect clinical study data given the current AE classification and data aggregation methods. Newly documented phase II study data are subject to the same environment. Therefore, when external controls are necessary, those should be derived from clinical trials to allow comparison of frequencies under the same conditions.

## Conclusion

The findings reported here support the novel approach of utilizing published trial data as external controls to interpret AE frequencies of single-arm studies. For fatigue and asthenia, we recommend that these two closely related terms be combined, results that exceed the highest-ever-reported frequencies in placebo arms should be flagged, and absolute values for reported AEs and total patient numbers should be used (Table III) for more precise assessments.

## Conflicts of Interest

HH is head of a pediatric palliative care team in Giessen, Germany and has no conflict of interest. BW is attending at Swedish Covenant hospital, and has no conflict of interest. JW is an employee of AbbVie pharmaceuticals Inc., and owns stocks. However, this project was not part of the employment, and the data interpretation reflect the personal opinion of the Authors, not the company.

## Authors' Contributions

BW: Concept, writing and figures. JW: Concept, data collection, analysis, writing and tables. HH: Concept, writing and tables.

## Acknowledgements

The Authors would like to thank Elisa Cerri and Guillermo Rivell for helpful discussions.

## References

- Taal W, Oosterkamp HM, Walenkamp AM, Dubbink HJ, Beerepoot LV, Hanse MC, Buter J, Honkoop AH, Boerman D, de Vos FY, Dinjens WN, Enting RH, Taphoorn MJ, van den Berkmortel FW, Jansen RL, Brandsma D, Bromberg JE, van Heuvel I, Vernhout RM, van der Holt B and van den Bent MJ: Single-agent bevacizumab or lomustine *versus* a combination of bevacizumab plus lomustine in patients with recurrent glioblastoma (BELOB trial): a randomised controlled phase 2 trial. *Lancet Oncol* 15(9): 943-953, 2014. PMID: 25035291. DOI: 10.1016/S1470-2045(14)70314-6
- Piha-Paul SA, Sachdev JC, Barve M, LoRusso P, Szmulewitz R, Patel SP, Lara PN Jr, Chen X, Hu B, Freise KJ, Modi D, Sood A, Hutti JE, Wolff J and O'Neil BH: First-in-human study of mivebresib (ABBV-075), an oral pan-inhibitor of bromodomain and extra terminal proteins, in patients with relapsed/refractory solid tumors. *Clin Cancer Res* 25(21): 6309-6319, 2019. PMID: 31420359. DOI: 10.1158/1078-0432.CCR-19-0578
- National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 2017. Available at: [https://ctep.cancer.gov/protocolDevelopment/electronic\\_applications/ctc.htm](https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm). [Last accessed on October 31st, 2021]
- Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET and Carbone PP: Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 5(6): 649-655, 1982. PMID: 7165009.
- Wolff BJ, Wolff MJ and Wolff JE: Headache reporting in oncology trials depends on the demographics of the study population. *In Vivo* 35(4): 1939-1943, 2021. PMID: 34182466. DOI: 10.21873/invivo.12460
- Wolff JE: Anemia in placebo arms of cancer studies. *Anticancer Res* 41(9): 4543-4548, 2021. PMID: 34475081. DOI: 10.21873/anticancer.15266
- Wolff B and Wolff J: Diarrhea in placebo arms of cancer studies. *Cancer Diagnosis & Prognosis* 1(5): 379-385, 2021. DOI: 10.21873/cdp.10050
- Chacón MR, Enrico DH, Burton J, Waisberg FD and Videla VM: Incidence of placebo adverse events in randomized clinical trials of targeted and immunotherapy cancer drugs in the adjuvant setting: a systematic review and meta-analysis. *JAMA Netw Open* 1(8): e185617, 2018. PMID: 30646278. DOI: 10.1001/jamanetworkopen.2018.5617
- Payne JM, Barton B, Ullrich NJ, Cantor A, Hearps SJ, Cutter G, Rosser T, Walsh KS, Gioia GA, Wolters PL, Tonsgard J, Schorry E, Viskochil D, Klesse L, Fisher M, Gutmann DH, Silva AJ, Hunter SJ, Rey-Cassery C, Cantor NL, Byars AW, Stavinoha PL, Ackerson JD, Armstrong CL, Isenberg J, O'Neil SH, Packer RJ, Korf B, Acosta MT, North KN and NF Clinical Trials Consortium: Randomized placebo-controlled study of lovastatin in children with neurofibromatosis type 1. *Neurology* 87(24): 2575-2584, 2016. PMID: 27956565. DOI: 10.1212/WNL.0000000000003435
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA and PRISMA-P Group: Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 4: 1, 2015. PMID: 25554246. DOI: 10.1186/2046-4053-4-1
- Niwa N, Kosaka T and Oya M: Placebo-induced severe diarrhea in a patient with metastatic castration-resistant prostate cancer participating in a double-blind, randomized, Phase III international trial. *Clin Genitourin Cancer* 14(2): e219-e220, 2016. PMID: 26774349. DOI: 10.1016/j.clgc.2015.12.027
- Day J, Yust-Katz S, Cachia D, Wefel J, Katz LH, Tremont I, Bulbeck H, Armstrong T and Rooney AG: Interventions for the management of fatigue in adults with a primary brain tumour. *Cochrane Database Syst Rev* 4: CD011376, 2016. PMID: 27074263. DOI: 10.1002/14651858.CD011376.pub2
- Meyer F, Freeman MP, Petrillo L, Barsky M, Galvan T, Kim S, Cohen L and Joffe H: Armodafinil for fatigue associated with menopause: an open-label trial. *Menopause* 23(2): 209-214, 2016. PMID: 26125538. DOI: 10.1097/GME.0000000000000505
- Poort H, Peters MEWJ, van der Graaf WTA, Nieuwkerk PT, van de Wouw AJ, Nijhuis-van der Sanden MWG, Bleijenberg G, Verhagen CAHHVM and Knoop H: Cognitive behavioral therapy or graded exercise therapy compared with usual care for severe fatigue in patients with advanced cancer during treatment: a randomized controlled trial. *Ann Oncol* 31(1): 115-122, 2020. PMID: 31912784. DOI: 10.1016/j.annonc.2019.09.002
- Heckler CE, Garland SN, Peoples AR, Perlis ML, Shayne M, Morrow GR, Kamen C, Hoefler J and Roscoe JA: Cognitive behavioral therapy for insomnia, but not armodafinil, improves fatigue in cancer survivors with insomnia: a randomized placebo-controlled trial. *Support Care Cancer* 24(5): 2059-2066, 2016. PMID: 26542272. DOI: 10.1007/s00520-015-2996-y
- Holt GB: Potential Simpson's paradox in multicenter study of intraperitoneal chemotherapy for ovarian cancer. *J Clin Oncol* 34(9): 1016, 2016. PMID: 26755513. DOI: 10.1200/JCO.2015.64.4542

Received October 5, 2021

Revised November 9, 2021

Accepted November 11, 2021